

COMMENT

Sungkyu Jung, Mark Foskey and J. S. Marron

The University of North Carolina at Chapel Hill

This paper proposes a novel PCA scheme for variables on a manifold. The authors tackled an interesting problem. In particular when the mass of a random quantity or a set of data is mostly aligned along a sufficiently long arc of a geodesic, the intrinsic and extrinsic mean no longer provide a useful notion of the center of mass due to the curvature of the manifold. The authors defined a new notion of mean, the Principal Component mean (PM), as the point where the first and second geodesic PCs meet. Regardless of the curvature of the manifold, this Geodesic PCA (GPCA) works well when most of the mass is along a geodesic.

The manifold data we encounter are typically S^2 (and $(S^2)^d$). In the data that we work with, we have noticed a tendency towards variation along smaller circles than geodesics, so we point out variations of the PCA method of this paper that better capture variation along a non-geodesic curve on manifolds.

The medial representations of shapes (m-reps, see e.g., Siddiqi and Pizer (2008), Fletcher, Lu, Pizer and Joshi (2004)) lie in direct products of S^2 , S^1 , R , and R^+ . In particular, in an example of randomly generated m-reps, that resemble human organs, from a simulator discussed in Jeong, Stough, Marron and Pizer (2008), we have observed S^2 -valued variables distributed along a small circle, which is not a geodesic (left panel of Figure 1.) The projection of the S^2 -valued data on geodesic PCs (center panel of Figure 1) results in a curvy form, where the first PC captures only approximately 95% of the variation. The method of PCA at IM (Fletcher et al. (2004)) also shows similar performance for this data set. We suggest an improvement, due to Gray, Geiser and Geiser (1980), that captures 100% of the variation of the first PC. This allows a non-geodesic curve to be a principal curve.

To illustrate this idea, let the simple manifold S^2 be our space. Let $x_1, \dots, x_n \in S^2$ be the data points. We wish to consider the principal modes of variation that are not necessarily along a geodesic, where the constraint of δ_1 being geodesics is relaxed to include all circles. Then a small circle δ_1 minimizing residual variance $\sum_i d(x_i, \delta_1)^2$, where d is the metric on S^2 , is defined to be the first mode of variation, i.e., the first PC. Furthermore, in the spirit of Definition 3.1, sequentially find δ_2 to minimize $\sum_i d(x_i, \delta_2)^2$ subject to orthogonally crossing δ_1 . This entails

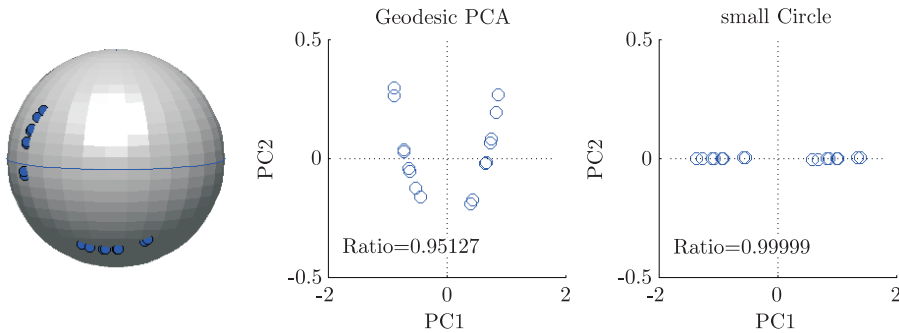


Figure 1. Sampling variation of the S^2 -valued variable in m -reps and projections on various PCs. (i) The data points are mostly along a small circle, which is not a geodesic. (ii) The method of GPCA: A set of data in (i) is projected to the tangent space at PM together with the two Geodesic PCs. (iii) Alternative method: The data set is projected on circles δ_1 and δ_2 . The proportion of variance explained by the first PC is shown as *ratios*.

that δ_2 is a geodesic. Then we can define a new center of mass as the intersection of δ_1 and δ_2 , say p , which minimizes $\sum_i d(x_i, p)^2$. The last plot in Figure 1 shows the projection of the data onto δ_1 and δ_2 , where the projection on δ_2 is not an orthogonal projection, but a projection along the direction of δ_1 .

These ideas, including extension to important manifolds such as $(S^2)^d$, will be further developed in Jung, Foskey and Marron (2009).

References

- Fletcher, P. T., Lu, C., Pizer, S. M. and Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. on Medical Imaging* **23**, 995-1005.
- Jeong, J.-Y., Stough, J. V., Marron, J. S. and Pizer, S. M. (2008). Conditional-mean initialization using neighboring objects in deformable model segmentation., *SPIE Medical Imaging*, February.
- Jung, S., Foskey, M. and Marron, J. S. (2009). Principal arc analysis on direct product manifolds, manuscript in progress.
- Gray, N. H., Geiser, P. A. and Geiser, J. R. (1980). On the least-squares fit of small and great circles to spherically projected orientation data. *Mathematical Geology* **12**, 173-184.
- Siddiqi, K. and Pizer, S. (2008). *Medial Representations: Mathematics, Algorithms and Applications*. Springer.

Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC27514, U.S.A.

E-mail: sungkyu@email.unc.edu

Department of Radiation Oncology, The University of North Carolina at Chapel Hill, Chapel Hill, NC27514, U.S.A.

E-mail: mark_foskey@unc.edu

Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC27514, U.S.A.

E-mail: marron@email.unc.edu

(Received December 2008; accepted February 2009)